SELECTION OF INFORMATIVE TEMPLATE IN HIERARCHICAL SPARSE METHOD

RAMADHAN A. M. ALSAIDI[1], HONG LI[2], HONGFENG LI[3], ROKAN KHAJI[4]

[1]Department of Mathematics, Science and Arts College, Aljouf university, Alqurayyat, Saudi Arabia
[2,3]School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China
[4]Department of Mathematics, College of Science, University of Diyala, Diyala,Iraq

abstract A dictionary of template is a key semantic component in a hierarchical sparse method (HSM). Since template selection attempts to eliminate redundant and irrelevant template, successful extract sufficient templates will lead to more discriminative ability of the HSM. In this paper, we present development of HSM by introduce a new template selection method based on the entropy concept. Its a way to indicate the total information the templates have. Algorithm is suggested that the template of more information should be picked and the others should be discarded. In this way, the proposed method provides HSM with better discriminative ability. Experimental results show that the introduced method achieves good performance in template selection with less computation. **keywords** Informative template; Entropy; Entropy of an image; Essential elements for hierarchical method

1 Introduction

The reliable and discriminative template is the main task for different template extraction methods to complete the recognition target. Several studies have been done to achieve this task. However, further research is needed to arrive at a more reliable and discriminative results. In fact, in this study, we looked at the recent hierarchical approach introduced by Smale[1]. It can be described as a multi-layer architecture, where each layer of the architecture is performing similar processes that are filtering and pooling operations. The key element in the hierarchical architecture is the dictionary of templates. The most trivial form of template selection proposed in [1, 2], the authors obtained templates by a uniform randomly subsampling an image set. Unfortunately, in this approach, the extracted features have high variance as the templates are resampled; especially when the number of templates is small, it leads to poor results. Furthermore, a large number of randomly selected templates do not always lead to better features, and the running time also increases. To solve this problem, Li et al proposed a new template selection method based on k-mean technique[3]. However, there are some disadvantages in the k-mean

algorithm that are listed below[4]:
First, It is sensitive to the selection of initial cluster center, that why the algorithm doesn't guarantee for optimal solution usually ended without global optimal solution, but suboptimal solution.
Second, there is no applicable evidence for the decision of the value of K (number of cluster to generate), and sensitive to initial value, for different initial values, there may be different clusters generated.
Third, This algorithm is disturbed easily by abnormal data; a few of abnormal data cause an extreme influence on the mean value.
Fourth, sometimes, the result of the cluster may lose balance.
Consequently, the selection template method based on k-mean is significantly affected by these disadvantages. Recently, Li et al. [5] used local coding approach to select the template in the first-layer. The second-layer templates are selected by making use the label information of the training images. Over the past decades, information theory has found a wide variety of applications in imaging and graphics.For example, to solve the enhancement task in imaging, Cheng et al. [6] proposed technique to perform image enhancement by transforming an image into a fuzzy domain with maximum fuzzy entropy. On the other hand, mutual information has been widely used in medical image registration since the early 1990s [7, 8]. Based on this, we propose a new technique to extract compact template sets with better discrimination ability. This method fills the gap by introduce a new technique to select a few templates among several templates in a context of hierarchical feature extraction based on entropy. In other words, it is a way to indicate the total information the templates have. It is recommended that the template of more information should be selected and the others should be discarded. This method uses low computational processes, which provide HSM to increase recognition rate within fewer computations and short time. The remaining parts of the paper are organized as follows: introducing the definition of the Entropy in section 2. Section 3 introduced the essential elements for proposed model and its related template selection method. We verified the effectiveness of the proposed method with extensive experiments on popular data set and compared it with the state-of-the-art algorithms in section 4. Finally, we present this work conclusion and propose future research trends in this topic in section 5.

2 Entropy

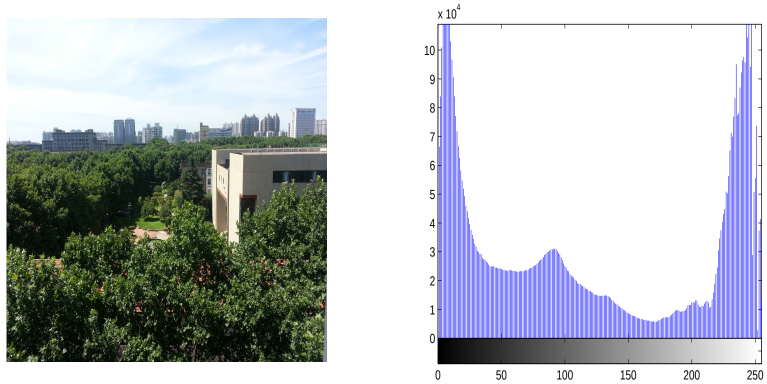
In 1948, Claude Shannon proposed A Mathematical Theory of Communication; it is widely accepted as the birth of Information Theory. Shannon used probability theory to model information sources. The information content, namely (Shannon) entropy of a discrete random variable X that has a probability distribution $px = (p_1, p_2, \cdots, p_n)$ is then defined as:

$$H(X) = H(px) \triangleq \sum_{i}^{n} p_i log \frac{1}{p_i} \tag{1}$$

The detail outlines, and treatment can be found in [9, 10, 11]

Entropy of an Image

Entropy concept arose within the study of physics of heat engines. It measures the amount of system disorder. Another way of expressing entropy is to consider the spread of states which a system can adopt. In the case of an image, these states correspond to the gray levels which the individual pixels can adopt. For example, in an 8-bit pixel there are 256 such states. If all such states are equally occupied, as they are in the case of an image which has been perfectly histogram equalized, the spread of states is a maximum, as the entropy of the image. On the other hand, if the image has been thresholded, only two states are occupied, the entropy is low. Indeed, the image is composed of pixels, and each pixel has a value, if the values of the pixels are equal; the entropy will be zero.



(a) Natural Image (b) Histogram

Figure 1: The natural image and its histogram.

Figure 1 shows a natural image and a histogram of the pixel intensity values. The normalized histogram can be an estimate of the underlying probability of pixel intensities, i.e., $p(i) = h_U(i)/N$, where $h_U(i)$ denotes the histogram entry of intensity value i in image U and N is the total number of pixels of U. Expending this model, we can compute the entropy of image as

$$H(U) = \sum_i h_U(i) log N / h_U(i) \tag{2}$$

3 Essential elements for proposed model

To proceed with the proposed method, it should be noted that, a hierarchical architecture in the proposed model consists of three layers of patches u, v and Sq in R^2, where $u \subset v \subset Sq$, which the paper assumes to be square. Practically, the ingredients needed to define the proposed framework are given as follows:

- A finite architecture defined by nested patches (for examples, subdomains of the square in R^2, see Fig.2)
- Transformations from a patch to the next larger one (see Fig. 3),
- At each layer we are given a function space denoted by:

$$I_u = \{f_1 : u \to [0, 1]\}$$
$$I_v = \{f_2 : v \to [0, 1]\}$$
$$I_{Sq} = \{f_3 : Sq \to [0, 1]\}$$

which defined on patches Sq, v and u respectively that we will be working with.

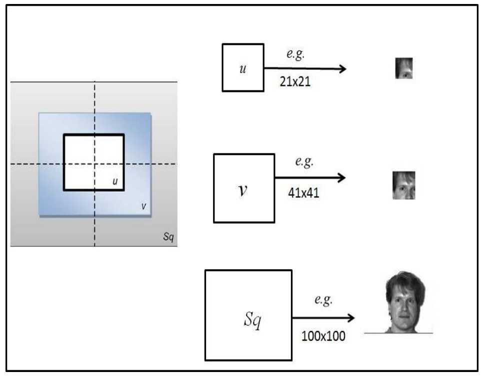


Figure 2: Nested patch domains

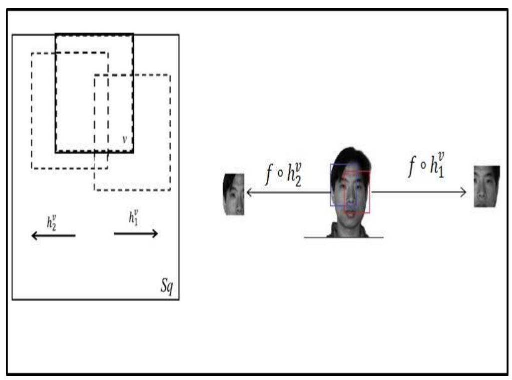


Figure 3: Translation operation

- Templates connecting the mathematical model to a real world setting.
  For a detailed treatment of the subject, the reader is referred to more dedicated works, such as [1, 12, 13, 2].

3.1 Selection of informative template

The goal of view selection is to automatically suggest interesting or optimal templates that maximize the amount of information received in the training set of a given image dataset.Good templates reveal essential information about the underlying data. Therefore, presenting them sooner to the viewers can improve both the speed and efficiency of data understanding. For example, in Fig. 4, we show three representative views of templates with different amounts of information revealed. Clearly, the rightmost one corresponds to the best view which reveals the maximum amount of information about the data by displaying the object in the least uncertain way. The proposed template selection method is presented

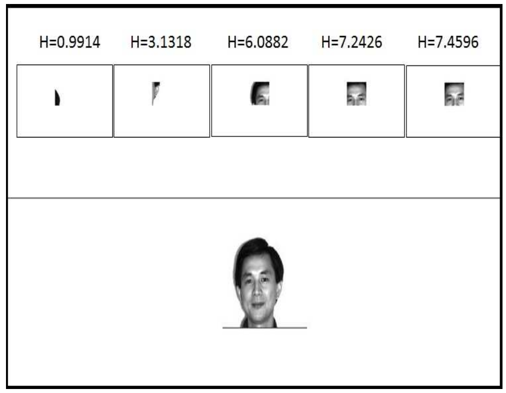


Figure 4: Five representative views of templates showing the increasing amount of information revealed about the object.

as follows: Let us set the initial template sets which randomly extracting image patches from an image set of size u and v. Note that, the size of u is relatively small because it contains only the basic elements of the objects. The result template of size u denoted by

$$T_u = \{t_1^u, t_1^u, \cdots, t_{q^u}^u\}$$

(where q^u is the cardinality of T_u and $T_u \in I_u$). On the other hand, the initial template of size v (where the size of v is large enough to contain more discriminative structure) denoted by

$$T_v = \{t_1^v, t_1^v, \cdots, t_{q^v}^v\}$$

(where q^v is the cardinality of T_v and $T_v \in I_v$)(To make the terms of template u and v clearer, an example is given in Fig. 2). Indeed, T_u and T_v have high variance as the template are resamples, particularly when q^u and q^v are small. On the other hand, a large number of q^u and q^v causes increase in the running time and do not always lead to better features. Construction of compact template sets with better discriminative ability is needed. Our task was to select the template sets P_u and P_v from the initial template sets T_u and T_v such that

$$P_u \subset T_u \subset I_u \; and \; P_v \subset T_v \subset I_v$$

3.2 Steps to select P_u from T_u:

1. Compute the value of entropy of each candidate template $T_i^u (i = 1, 2, \cdots, q^u)$.
2. Rank T_u in ascending order according to their entropy.

3. Select the last T_u candidate template in the ranked set to construct P_u.

In the same way we can select P_v from T_v.

Algorithm 1 Construction of P_u

Input: Full template set $T_u = \left\{t_1^u, t_1^u, \cdots, t_{q^u}^u\right\}$ and number of template d
Output: P_u

1. count=1
2. $P_u = \phi$
3. $t_k = argmax_{t_k \in T_u} H\left(t_k\right)$
4. $P_u = P_u \cup t_k$
5. $T_u = T_u \setminus t_k$ $and\ count = count + 1$
6. $if\ count < d\ go\ to\ 3$
7. $else$
8. $P_u = P_u \cup \phi$
9. $end\ if$

Finally, it should be noted that only at the stage of choosing template there is a change and the rest of the steps are the same as [12, 13].

4 Experimental results and analysis

We will demonstrate the proposed method against four state-of-the-art methods on Yale face database.The Yale Face dataset (size 6.4MB) consists of 165 grayscale images in GIF format of 15 classes. There are 11 images per subject, one per different facial expression or configuration: left-light,center-light, w/glasses, normal, w/no glasses, happy , right-light, sad, sleepy, wink and surprised (see fig. 5). In this experiment, we randomly selected one image from each class to construct template sets and used the rest for training and testing. In the experiments,we compare the following algorithm for the development method:
1: Neural Response (NR) [1, 3].
2: Hierarchical Model And X (HMAX) [14, 15]

Table 1: Comparison results with Yale face dataset

Methods	**Accuracy**
HMAX Model	85.63
NR	85.21
Developed	**89.07**

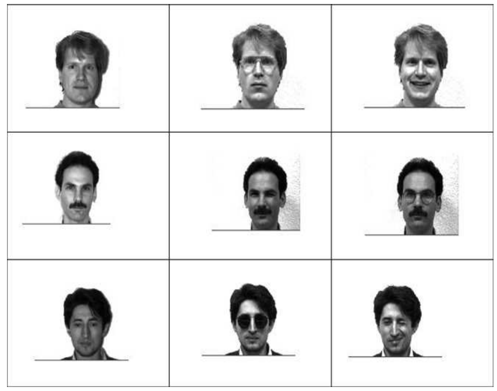


Figure 5: Nine instances from the set of training examples.

Nevertheless, the performance of the proposed method is significantly affected by the following parameters: patch size (u,v), size of the available training set (Tr), number of elements in each linear combination (L) and the cardinality. After carefully scrutinizing the above experiments, we find the optimal parameters values used to achieve this accuracy are: $L = 5$, $u = 21$, $v = 30$. In addition for the cardinality, the proposed implementation selected 165 templates among 1155 for first layer and selected 195 templates among 1020 for second layer. Recognition accuracies were averaged over five random splits. On the other hand, all the image patches that are used for constructing template sets come from the picked images. These picked images are not included in the test or training sets[1]. The classification accuracies are identified by the pcaForSVM classifier which combined Support Vector Machine (SVM) and Principle Component Analysis (PCA) to pretreatment and reduce dimension of feature data. We implement all processing in grayscale even, when color images are available.

5 Conclusions

We have presented in this article a simple scheme for binary template selection in a context of feature extraction based on entropy. It is experimentally shown that the proposed algorithm can be employed to selected an effective template. In addition, this method uses low computational processes, which provide HSM to increase recognition rate within fewer computations and short time. The fundamental challenge for proposed method, consider a set of template containing several templates almost identical. Picking one of them makes all the other ones of this group useless during the rest of the computation. Hence, how to adopt our method not to select a template similar to already picked ones? For feature work also, we suggest to benefit from various techniques in feature selection approach and attempt to use it to select the template.

References

[1] Smale S, Rosasco L, Bouvrie J, et al. Mathematics of the neural response. Foundations of Computational Mathematics, 2010, 10(1):67–91.

[2] Bouvrie J, Poggio T, Rosasco L, et al. Generalization and properties of the neural response. 2010.

[3] Li H, Wei Y, Li L, et al. Similarity learning for object recognition based on derived kernel. Neurocomputing, 2012, 83:110–120.

[4] Hartigan J A, Wong M A. Algorithm AS 136: A k-means clustering algorithm. Applied statistics, 1979, pages 100–108.

[5] Li H, Wei Y, Li L, et al. Hierarchical feature extraction with local neural response for image recognition. Cybernetics, IEEE Transactions on, 2013, 43(2):412–424.

[6] Cheng H, Chen Y H, Sun Y. A novel fuzzy entropy approach to image enhancement and thresholding. Signal Processing, 1999, 75(3):277–301.

[7] Viola P, Wells III W M. Alignment by maximization of mutual information. International journal of computer vision, 1997, 24(2):137–154.

[8] Pluim J P. Mutual information based registration of medical images. Ponsen & Looijen, 2000, pages p22–p25.

[9] Cover T M, Thomas J A. Elements of information theory. John Wiley & Sons, 2012.

[10] Shannon C E. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review, 2001, 5(1):3–55.

[11] Sabuncu M R. Entropy-based image registration: [PhD Dissertation]. Citeseer, 2004.

[12] Alsaidi R A M, Li H, Wei Y, et al. Hierarchical sparse method with applications In vision and speech recognition. International Journal of Wavelets, Multiresolution and Information Processing, 2013, 11(02).

[13] Alsaidi R A M, Li H, Wei Y, et al. Improved Hierarchical Sparse Method with application to Offline handwritten Arabic character recognition. International Journal of Scientific, Engineering Research, 2014, 5(03).

[14] Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007, 29(3):411–426.

[15] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature neuroscience, 1999, 2(11):1019–1025.